

CC protein. The method comprises providing a sample of a body fluid and/or  
 CC tissue of a patient and determining a level of the aberrant protein  
 CC and/or its detectable part or the ratio between the two in the body fluid  
 CC and/or tissue. The method enables diagnosing the disease before the  
 CC patient exhibits clinical symptoms. The method is useful for diagnosing  
 CC age-related diseases, preferably Alzheimer's disease, Down's syndrome or  
 CC other age related diseases e.g. cancer; neurodegenerative diseases, such  
 CC as frontal lobe dementia, progressive supranuclear palsy; and other  
 CC diseases with abundant tau-positive filamentous lesions, Parkinson's  
 CC disease, amyotrophic lateral sclerosis, Huntington's disease,  
 CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body  
 CC diseases associated with ubiquitin, diabetes mellitus type II and other  
 CC degenerative diseases such as cardiovascular diseases and rheumatoid  
 CC arthritis. A number of other diseases which can be diagnosed are given in  
 CC the specification. The present sequence is AMY 6 peptide, which was  
 CC coupled to thyroglobulin by glutaraldehyde, and used to immunise  
 CC rabbits for antibody production in the present invention. The AMY 6  
 CC peptide is a C-terminal fragment of the full-length beta-amyloid  
 CC precursor protein (APP+1; mature APP and its signal peptide) protein.  
 XX  
 SQ Sequence 19 AA:

Query Match 83.3%; Score 82.5; DB 22; Length 19;  
 Best Local Similarity 94.7%; Pred. No. 1.6e-07;  
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 NVPCHERMGRGRTSSKELA 19  
 DB 2 NVP-HERMGRGRTSSKELA 19  
 ||||| ||||| ||||| ||||| |||||

RESULT 8  
 AAB99215  
 ID AAB99215 standard; peptide; 14 AA.

AC AAB99215;

DT 06-SEP-2001 (first entry)

DE AMY5 peptide.

XX AMY5; disease detection; age-related disease; Alzheimer's disease;  
 KW Down's syndrome; cancer; neurodegenerative disease; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Huntington's disease;  
 KW spinocerebellar ataxia-3; multiple sclerosis; diabetes mellitus type II;  
 KW degenerative disease; cardiovascular disease; rheumatoid arthritis;  
 KW beta-amyloid precursor protein; APP.

XX Unidentified.

XX WO200140804-A2.

XX 07-JUN-2001.

XX 04-DEC-2000; 2000WO-NL00887.

XX 03-DEC-1999; 99EP-0204140.

XX (NEWV-) KONINK NEDERLANDSE AKAD VAN WETENSCHAPPE.

XX Hol EM, Van Leeuwen FW;

XX WPI; 2001-397965/42.

XX Detecting a disease or pathological condition, associated with  
 PT secretion of aberrant protein e.g. age-related diseases, by determining  
 PT level of aberrant protein and/or its detectable part in body fluid  
 PT and/or tissue

XX Disclosure; Page 24; 36pp; English.

XX The present invention relates to a method for detecting a disease or  
 CC pathological condition associated with molecular misreading of coding

CC sequences in the genome and/or associated  
 CC protein. The method comprises providing  
 CC tissue of a patient and determining a le  
 CC and/or its detectable part or the ratio be  
 CC and/or tissue. The method enables diagnosing  
 CC patient exhibits clinical symptoms. The metho  
 CC age-related diseases, preferably Alzheimer's d  
 CC as frontal lobe dementia, progressive supranuclea  
 CC diseases with abundant tau-positive filamentous les  
 CC disease, amyotrophic lateral sclerosis, Huntington's  
 CC spinocerebellar ataxia-3, multiple sclerosis; other inc  
 CC diseases associated with ubiquitin, diabetes mellitus ty  
 CC degenerative diseases such as cardiovascular diseases and rheumatoid  
 CC arthritis. A number of other diseases which can be diagnosed are given in  
 CC the specification. The present sequence is AMY 5 peptide, which was  
 CC coupled to thyroglobulin by glutaraldehyde, and used to immunise  
 CC rabbits for antibody production in the present invention. The AMY 5  
 CC peptide is a C-terminal fragment of the full-length beta-amyloid  
 CC precursor protein (APP+1; mature APP and its signal peptide) protein.  
 XX  
 SQ Sequence 14 AA:

Query Match 63.6%; Score 63; DB 22; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.00033;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMGRGRTSSKELA 19  
 ||||| ||||| |||||

DB 2 RMGRGRTSSKELA 14

RESULT 9  
 AAY21522

ID AAY21522 standard; Protein; 28 AA.

AC AAY21522;

DT 22-JUL-1999 (first entry)

DE Human beta-APP exon 10 protein fragment.

XX Human: beta-amyloid precursor protein; beta-APP; diagnosis: cancer;  
 KW frameshift mutation; age-related disease; neurodegenerative disorder;  
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;  
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;  
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;  
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;  
 KW neurofilament-F; presenilin 1; presenilin 2; cellular tumour antigen;  
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;  
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;  
 KW high mobility group protein-C; neuroendocrine specific protein A.

OS Synthetic.

OS Homo sapiens.

XX WO9845322-A2.

XX 15-OCT-1998.

XX 02-APR-1998; 98WO-IB00705.

XX 10-APR-1997; 97US-0043163.

XX (UKYT-) RIJKSUNIV UTRECHT.

XX (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.

XX (UYRO-) UNIV ROTTERDAM ERASMUS.

XX Burbach JPH, Grosveld FG, Van Leeuwen FW;

XX WPI; 1998-609901/51.

XX N-PSDB; AAX75772.

PT Diagnosing disease by detecting frameshift mutations in RNA or  
 PT corresponding protein mutations - used to diagnose cancer and  
 PI neurological diseases, particularly Alzheimer's disease, and also  
 PI for treatment and prevention with specific ribozymes or wild-type  
 PT RNA

PS Disclosure: Figure 20; 258pp; English.

XX This invention describes a novel method for the diagnosis of a disease  
 CC caused by, or associated with, an RNA molecule that has a frameshift  
 CC mutation. The method is used to diagnose age-related diseases, especially  
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's  
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,  
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II  
 CC and many others listed) or susceptibility to these disorders. The method  
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,  
 CC at an early stage. It is based on the observation that disease may be  
 CC caused by mutations in RNA rather than DNA. The invention describes the  
 CC use of neuronal system RNA molecules, specifically proteins including  
 CC beta-amyloid precursor protein (beta-APP), the microtubule associated  
 CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule  
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,  
 CC neurofilament-F, presenilin 1, presenilin 2, glial fibrillary acidic  
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma  
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group  
 CC protein-C (HMGP-C) and neuroendocrine specific protein A.

XX Sequence 28 AA;

Query Match 63.6%; Score 63; DB 19; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 0.0007;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RMGRGRTSSKELA 19  
 |||||  
 Db 16 RMGRGRTSSKELA 28

RESULT 10  
 AAY56394

ID AAY56394 standard; peptide; 12 AA.

XX AAY56394;

DT 15-FEB-2000 (first entry)

DE Mutant beta-APP protein peptide sequence SEQ ID NO:4.

KW Mutant: beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;  
 KW Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;  
 KW gene therapy.

XX Synthetic.

OS Homo sapiens.

XX WO9958564-A1.

PN 18-NOV-1999.

PD 30-APR-1999; 99WO-NO00141.

XX 08-MAY-1998; 98NO-0002098.

XX (NHYP ) NORSE HYDRO AS.

PI Gaudernack G, Eriksen JA, Moller M;

XX WPI; 2000-039070/03.

XX Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,  
 PT useful for treating Alzheimer's disease and Down syndrome.

PS Claim 10; Page 27; 33pp; English.

XX The present invention describes frameshift mutant beta-amyloid precursor  
 CC peptides (beta-APP) (AAY56391 to AAY56394) and mutant ubiquitin-B  
 CC (Ubi-B) (AAY56395 to AAY56400) associated with Alzheimer's disease and  
 CC Down syndrome eliciting T cellular immunity. The peptides may be used as  
 CC a vaccine for Alzheimer's disease and Down syndrome. The vaccination is  
 CC sufficient to induce specific T-cell immunity to mutant beta-APP and/or  
 CC Ubi-B peptides that are associated with Alzheimer's disease or Down  
 CC syndrome. The patient may be stimulated in vivo or ex vivo with the  
 CC peptides. The peptides and DNA encoding the peptides can also be used for  
 CC the treatment or prophylaxis of Alzheimer's disease or the treatment of  
 CC Down syndrome.

XX Sequence 12 AA;

Query Match 58.6%; Score 58; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.0021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGRGRTSSKELA 19  
 |||||  
 Db 1 MGRGRTSSKELA 12

RESULT 11

AAY56393

ID AAY56393 standard; peptide; 17 AA.

XX AAY56393;

DT 15-FEB-2000 (first entry)

DE Mutant beta-APP protein peptide sequence SEQ ID NO:3.

KW Mutant: beta-amyloid precursor protein; beta-APP; ubiquitin-B; vaccine;  
 KW Alzheimer's disease; Down syndrome; T cellular immunity; neuroprotective;  
 KW gene therapy.

XX Synthetic.

OS Homo sapiens.

XX WO9958564-A1.

PD 18-NOV-1999.

XX 30-APR-1999; 99WO-NO00141.

XX 08-MAY-1998; 98NO-0002098.

XX (NHYP ) NORSE HYDRO AS.

PI Gaudernack G, Eriksen JA, Moller M;

XX WPI; 2000-039070/03.

XX Frameshift mutant beta-amyloid precursor and ubiquitin-B peptides,  
 PT useful for treating Alzheimer's disease and Down syndrome.

PS Claim 10; Page 27; 33pp; English.

XX The present invention describes frameshift mutant beta-amyloid precursor  
 CC peptides (beta-APP) (AAY56391 to AAY56394) and mutant ubiquitin-B  
 CC (Ubi-B) (AAY56395 to AAY56400) associated with Alzheimer's disease and  
 CC Down syndrome eliciting T cellular immunity. The peptides may be used as  
 CC a vaccine for Alzheimer's disease and Down syndrome. The vaccination is  
 CC sufficient to induce specific T-cell immunity to mutant beta-APP and/or  
 CC Ubi-B peptides that are associated with Alzheimer's disease or Down  
 CC syndrome. The patient may be stimulated in vivo or ex vivo with the  
 CC peptides. The peptides and DNA encoding the peptides can also be used for  
 CC the treatment or prophylaxis of Alzheimer's disease or the treatment of  
 CC Down syndrome.

XX Sequence 17 AA;

Query Match 52.5%; Score 52; DB 21; Length 17;  
 Best Local Similarity 100.0%; Pred. NO. 0.036;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGRMG 9  
 | | | | | | | |  
 DB 9 NVPGRMG 17

RESULT 12  
 AAJ46738  
 ID AAU46738 standard; Protein; 57 AA.  
 XX  
 AC AAU46738;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #7634.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WQ200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI: 2001-616774/71.  
 DR N-PSDB; AAS59535.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 7933; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in:  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 57 AA;

Query Match 51.5%; Score 51; DB 22; Length 57;  
 Best Local Similarity 81.8%; Pred. NO. 0.2;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPGHERMGRGR 12  
 | | | | | | | | | |  
 DB 47 VPGHRETGRGR 57

RESULT 13  
 AAU66292  
 ID AAU66292 standard; Protein; 57 AA.  
 XX  
 AC AAU66292;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #27188.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WQ200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI: 2001-616774/71.  
 DR N-PSDB; AAS59726.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 27487; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in:  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 57 AA;

```

Query Match      51.5%; Score 51; DB 22; Length 57;
Best Local Similarity 81.8%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 VPGHERMGRGR 12
Db 47 VPGHRTGGR 57

RESULT 14
ABB59130
ID ABB59130 standard; Protein: 491 AA.
XX
AC ABB59130;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4182.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NV.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL03233.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4182; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 491 AA:

Query Match      51.0%; Score 50.5; DB 22; Length 491;
Best Local Similarity 48.0%; Pred. No. 2.6;
Matches 12; Conservative 3; Mismatches 3; Indels 7; Gaps 1;

Qy 2 VPGHERMGRGRSS-----KELA 19
Db 171 LPsherLGLGRSLQSFTIELTKELA 195

RESULT 15
ABB53619
ID ABB53619 standard; Protein: 72 AA.

```

```

XX ABB53619;
AC
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein ydBC.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Roblotinc A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification of Lactococcus
PT lactis and related species -
XX
PS Claim 6; SEQ ID NO 321; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 72 AA:

Query Match      50.5%; Score 50; DB 23; Length 72;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSKE 17
Db 46 PDHEKMGKGITLSEF 60

Search completed: October 1, 2003, 09:38:25
Job time : 84 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:36:59 : Search time 17 Seconds  
(without alignments)  
47.289 Million cell updates/sec

Title: US-09-674-913a-1

Perfect score: 99

Sequence: 1 NVPGHMRGRGRTSSKELA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_AA: \*
- 1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep: \*
  - 2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep: \*
  - 3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep: \*
  - 4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep: \*
  - 5: /cgn2.6/ptodata/1/1aa/PctUS\_COMB.pep: \*
  - 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	100.0	25	2	US-08-726-306A-61
2	99	100.0	25	2	US-08-726-306A-87
3	47	47.5	10	2	US-08-726-306A-1
4	47	47.5	87	4	US-09-107-532A-5378
5	46	46.5	26	1	US-07-942-245-321
6	46	46.5	139	4	US-09-252-991A-23482
7	46	46.5	553	4	US-09-252-991A-17429
8	45.5	46.0	537	4	US-09-134-001C-4091
9	45	45.5	496	4	US-09-252-991A-22592
10	45	45.5	582	4	US-09-252-991A-25313
11	45	45.5	940	4	US-09-252-991A-24639
12	44	44.4	266	4	US-09-252-991A-32512
13	44	44.4	341	4	US-09-252-991A-27955
14	44	44.4	417	4	US-09-252-991A-16986
15	44	44.4	530	4	US-09-252-991A-30947
16	44	44.4	656	4	US-09-252-991A-22270
17	43	43.4	119	2	US-08-581-528A-7
18	43	43.4	119	5	PCT-US94-07799-7
19	43	43.4	161	2	US-08-581-528A-6
20	43	43.4	161	5	PCT-US94-07799-6
21	43	43.4	192	4	US-09-252-991A-8164
22	43	43.4	223	4	US-09-252-991A-17072
23	43	43.4	460	4	US-09-252-991A-26877
24	42	42.4	129	1	US-08-360-914B-15
25	42	42.4	129	1	US-08-741-589A-13
26	42	42.4	293	4	US-09-252-991A-29137
27	42	42.4	294	1	US-08-362-670B-2

28	42	42.4	294	3	US-08-333-576C-2	Sequence 2, Appli
29	42	42.4	294	3	US-08-808-324-2	Sequence 2, Appli
30	42	42.4	294	5	PCT-US94-14030A-2	Sequence 2, Appli
31	42	42.4	388	1	US-08-362-670B-34	Sequence 34, Appl
32	42	42.4	388	3	US-08-333-576C-34	Sequence 34, Appl
33	42	42.4	388	3	US-08-808-324-34	Sequence 34, Appl
34	42	42.4	388	5	PCT-US94-14030A-34	Sequence 34, Appl
35	42	42.4	411	1	US-08-362-670B-28	Sequence 28, Appl
36	42	42.4	411	3	US-08-333-576C-28	Sequence 28, Appl
37	42	42.4	411	3	US-08-808-324-28	Sequence 28, Appl
38	42	42.4	411	5	PCT-US94-14030A-28	Sequence 28, Appl
39	42	42.4	418	4	US-09-648-281-12	Sequence 12, Appl
40	42	42.4	591	4	US-09-252-991A-31286	Sequence 31286, A
41	42	42.4	592	4	US-09-252-991A-18124	Sequence 18124, A
42	42	42.4	735	4	US-09-252-991A-17053	Sequence 17053, A
43	41.5	41.9	684	4	US-09-252-991A-30736	Sequence 30736, A
44	41	41.4	123	4	US-09-732-210-1370	Sequence 1370, Ap
45	41	41.4	415	4	US-09-252-991A-27669	Sequence 27669, A

## ALIGNMENTS

RESULT 1  
US-08-726-306A-61  
: Sequence 61, Application US/08726306A  
: Patent No. 5958884  
: GENERAL INFORMATION:  
: APPLICANT: van Leeuwen, Frederik Willem  
: APPLICANT: Burbach, Johannes Peter Henri  
: APPLICANT: Grosveld, Franklin G.  
: TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
: NUMBER OF SEQUENCES: 189  
: CORRESPONDENCE ADDRESS:  
: ADDRESS: Banner & Witcoff, Ltd.  
: STREET: 1 Financial Center  
: CITY: Boston  
: STATE: MA  
: COUNTRY: US  
: ZIP: 02111  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
: COMPUTER: IBM PC Compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Wordperfect 6.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/726,306A  
: FILING DATE: 02-Oct-1996  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GB 95/20080.4  
: FILING DATE: 02-Oct-1995  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 60/009,832  
: FILING DATE: 01-Jan-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Williams, Ph.D., Kathleen M.  
: REGISTRATION NUMBER: 34,380  
: REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 345-9100  
: TELEFAX: (617) 345-9111  
: INFORMATION FOR SEQ ID NO: 61:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 25 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: unknown  
: MOLECULE TYPE: protein  
US-08-726-306A-61

Query Match 100.0%; Score 99: DB 2: Length 25;  
Best Local Similarity 100.0%; Pred. No. 5.1e-11;  
Matches 19: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYPGHERMGRGTSSKELA 19  
 DB 7 NYPGHERMGRGTSSKELA 25

## RESULT 2

US-08-726-306A-87  
 ; Sequence 87, Application US/08726306A  
 ; Patent No. 5958684  
 ; GENERAL INFORMATION:  
 ; APPLICANT: van Leeuwen, Frederik Willem  
 ; APPLICANT: Burbach, Johannes Peter Henri  
 ; APPLICANT: Grosveld, Franklin G.  
 ; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
 ; NUMBER OF SEQUENCES: 189  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Banner & Witcoff, Ltd.  
 ; STREET: 1 Financial Center  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02111

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WordPerfect 6.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/726.306A  
 FILING DATE: 02-Oct-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 95/20080.4  
 FILING DATE: 02-Oct-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/009,832  
 FILING DATE: 01-Jan-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Williams, Ph.D., Kathleen M.  
 REGISTRATION NUMBER: 34,380  
 REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 345-9100  
 TELEFAX: (617) 345-9111  
 INFORMATION FOR SEQ ID NO: 87:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 25 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-726-306A-87

Query Match 100.08; Score 99; DB 2; Length 25;  
 Best Local Similarity 100.08; Pred. No. 5,1e-11;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYPGHERMGRGTSSKELA 19  
 DB 7 NYPGHERMGRGTSSKELA 25

## RESULT 3

US-08-726-306A-1  
 ; Sequence 1, Application US/08726306A  
 ; Patent No. 5958684  
 ; GENERAL INFORMATION:  
 ; APPLICANT: van Leeuwen, Frederik Willem  
 ; APPLICANT: Burbach, Johannes Peter Henri  
 ; APPLICANT: Grosveld, Franklin G.  
 ; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
 ; NUMBER OF SEQUENCES: 189  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.  
 STREET: 1 Financial Center  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US  
 ZIP: 02111

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WordPerfect 6.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/726.306A  
 FILING DATE: 02-Oct-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 95/20080.4  
 FILING DATE: 02-Oct-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/009,832  
 FILING DATE: 01-Jan-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Williams, Ph.D., Kathleen M.  
 REGISTRATION NUMBER: 34,380  
 REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 345-9100  
 TELEFAX: (617) 345-9111  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-726-306A-1

Query Match 47.5%; Score 47; DB 2; Length 10;  
 Best Local Similarity 100.08; Pred. No. 0.032;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RGRITSSKELA 19  
 DB 1 RGRITSSKELA 10

## RESULT 4

US-09-107-532A-5378  
 ; Sequence 5378, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A. Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571

```
;
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneko
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5067
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8) LOCATION 1...87
; SEQUENCE DESCRIPTION: SEQ ID NO: 5378:
US-09-107-532A-5378

Query Match 47.5%; Score 47; DB 4; Length 87;
Best Local Similarity 53.3%; Pred. No. 0.41;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 PGHERMGRGRTSSKE 17
DB 61 PDHEKMGKGLTLTNE 75

RESULT 5
US-07-942-245-321
; Sequence 321, Application US/07942245
; Patent No. 5639641
; GENERAL INFORMATION:
; APPLICANT: PEDERSEN, Jan T.
; APPLICANT: REES, Stephen M.J.
; APPLICANT: REES, Anthony R.
; APPLICANT: ROGUSKA, Michael A.
; APPLICANT: GUILD, Braydon C.
; TITLE OF INVENTION: SURFACE RESIDUE VENEERING OF RODENT
; NUMBER OF SEQUENCES: 522
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: HP 9000/700 Workstation
; OPERATING SYSTEM: UNIX
; SOFTWARE: In house
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/942,245
; FILING DATE: 09-SEP-1992
; CLASSIFICATION: 530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 321:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-942-245-321

; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneko
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5067
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8) LOCATION 1...87
; SEQUENCE DESCRIPTION: SEQ ID NO: 5378:
US-09-107-532A-5378

Query Match 46.5%; Score 46; DB 1; Length 26;
Best Local Similarity 66.7%; Pred. No. 0.15;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PGHERMGRGRTS 14
DB 13 PGHEKKGRSSS 24

RESULT 6
US-09-252-991A-23482
; Sequence 23482, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/034,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23482
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23482

Query Match 46.5%; Score 46; DB 4; Length 139;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 VPGHERMGRGRTSSKELA 19
DB 53 VPGRAGAGRGRTTARSRA 70

RESULT 7
US-09-252-991A-17429
; Sequence 17429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/034,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17429
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17429

Query Match 46.5%; Score 46; DB 4; Length 553;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PGHERMGRGRT 13
DB 1 PDHQRHGRGRT 11

RESULT 8
US-09-134-001C-4091
```

```
; Sequence 4091, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4091
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4091

Query Match 46.0%; Score 45.5; DB 4; Length 537;
Best Local Similarity 55.0%; Pred. No. 6.3;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 NVPGH-ERMGRGRTSSKLA 19
I:|||||II:|||||
Db 186 NLPGHVAMGVDSQSTJLJA 205

RESULT 9
US-09-252-991A-22592
; Sequence 22592, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22592
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22592

Query Match 45.5%; Score 45; DB 4; Length 496;
Best Local Similarity 36.7%; Pred. No. 7.1;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

Qy 2 VPGHMR-----GRGT 13
I:|||||
Db 428 LPGHMRARLSKLTIDLVDYLRPGRGT 457

RESULT 10
US-09-252-991A-25313
; Sequence 25313, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25313
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25313

Query Match 45.5%; Score 45; DB 4; Length 582;
Best Local Similarity 50.0%; Pred. No. 8.6;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSK 16
I:|||||
Db 24 PGHRRAGRGRRGR 37

RESULT 11
US-09-252-991A-24639
; Sequence 24639, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24639
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24639

Query Match 45.5%; Score 45; DB 4; Length 940;
Best Local Similarity 52.9%; Pred. No. 15;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSKLA 19
I:|||||
Db 54 PGHFRFRGRSAAPELA 70

RESULT 12
US-09-252-991A-32512
; Sequence 32512, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32512
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32512

Query Match 44.4%; Score 44; DB 4; Length 266;
Best Local Similarity 72.7%; Pred. No. 5.2;
```



```
Matches      8;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;

QY      3 PGHERMGRGRT 13
      || | | | |
DB      245 PGHRRGRGRT 255

RESULT 13
US-09-252-991A-27955
; Sequence 27955, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27955
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27955

Query Match      44.4%; Score 44; DB 4; Length 341;
Best Local Similarity 53.8%; Pred. No. 6.9;
Matches      7;  Conservative      4;  Mismatches      2;  Indels      0;  Gaps      0;

QY      3 PGHERMGRGRTSS 15
      || | | | |
DB      87 PGADRLGGRTGA 99

RESULT 14
US-09-252-991A-16986
; Sequence 16986, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16986
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16986

Query Match      44.4%; Score 44; DB 4; Length 417;
Best Local Similarity 87.5%; Pred. No. 8.7;
Matches      7;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      3 PGHERMGR 10
      || | | | |
DB      42 PGHORMGR 49

RESULT 15
US-09-252-991A-30947
; Sequence 30947, Application US/09252991A
; Patent No. 6551795
```

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30947
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30947
```

```
Query Match      44.4%; Score 44; DB 4; Length 530;
Best Local Similarity 72.7%; Pred. No. 12;
Matches      8;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;
```

```
QY      3 PGHERMGRGRT 13
      || | | | |
DB      70 PGHPRAGRGT 80
```

```
Search completed: October 1, 2003, 09:41:48
Job time : 19 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:40:50 ; Search time 64 seconds  
(without alignments)  
46.969 Million cell updates/sec

Title: US-09-674-913A-1

Perfect score: 99  
Sequence: 1 NVPCHRMGRGRTSSKELA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 156212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	100.0	19	12 US-09-674-913A-1	Sequence 1, Appli
2	99	100.0	27	12 US-09-674-913A-2	Sequence 2, Appli
3	58	58.6	12	12 US-09-674-913A-4	Sequence 4, Appli
4	52	52.5	17	12 US-09-674-913A-3	Sequence 3, Appli
5	45	45.5	31	9 US-09-674-913A-1	Sequence 39237, A
6	45	45.5	70	11 US-09-674-913A-1	Sequence 39237, A
7	45	45.5	124	12 US-09-674-913A-1	Sequence 4632, Ap
8	43	43.4	480	15 US-09-674-913A-1	Sequence 70, Appl
9	43	43.4	551	14 US-10-156-761-11800	Sequence 11800, A
10	42	42.4	294	10 US-10-005-057A-2	Sequence 2, Appli
11	42	42.4	353	15 US-09-945-182-2	Sequence 2, Appli
12	42	42.4	388	10 US-09-945-182-34	Sequence 34, Appl
13	42	42.4	411	10 US-09-945-182-28	Sequence 28, Appl
14	42	42.4	418	12 US-10-141-541-2	Sequence 2, Appli
15	42	42.4	450	15 US-10-188-246-12	Sequence 12, Appl

16	42	42.4	1285	12 US-10-141-541-4	Sequence 4, Appli
17	41	41.4	15	12 US-09-674-913A-5	Sequence 5, Appli
18	40.5	40.9	218	11 US-09-975-719-15	Sequence 15, Appl
19	40.5	40.9	498	15 US-10-156-761-14582	Sequence 14582, A
20	40.5	40.9	502	15 US-10-156-761-11689	Sequence 11689, A
21	40.5	40.9	510	15 US-10-156-761-11980	Sequence 11980, A
22	40	40.4	116	9 US-09-867-550-1898	Sequence 1898, Ap
23	40	40.4	283	15 US-10-270-875-52	Sequence 52, Appl
24	40	40.4	283	15 US-10-270-878-52	Sequence 52, Appl
25	40	40.4	283	15 US-10-270-786-52	Sequence 52, Appl
26	40	40.4	283	15 US-10-270-710-52	Sequence 52, Appl
27	40	40.4	283	15 US-10-270-859-52	Sequence 52, Appl
28	40	40.4	322	12 US-10-238-075-330	Sequence 330, App
29	40	40.4	322	12 US-10-238-075-330	Sequence 10628, A
30	40	40.4	516	15 US-10-156-761-10628	Sequence 607, App
31	39	39.4	120	11 US-09-974-879-607	Sequence 608, App
32	39	39.4	120	11 US-09-305-736-608	Sequence 63, Appl
33	39	39.4	202	16 US-10-090-170-63	Sequence 63, Appl
34	39	39.4	240	10 US-09-945-182-30	Sequence 5773, Ap
35	39	39.4	247	10 US-09-738-626-5773	Sequence 97, Appl
36	39	39.4	291	12 US-10-074-511-97	Sequence 14366, A
37	39	39.4	368	15 US-10-156-761-14366	Sequence 13401, A
38	39	39.4	443	15 US-10-156-761-13401	Sequence 2, Appli
39	39	39.4	538	15 US-10-013-803A-2	Sequence 6, Appli
40	39	39.4	1312	12 US-10-141-541-6	Sequence 4, Appli
41	39	39.4	1745	12 US-09-793-061-4	Sequence 2, Appli
42	39	39.4	2374	11 US-09-383-894-2	Sequence 4, Appli
43	39	39.4	2425	11 US-09-383-894-4	Sequence 43384, A
44	38	38.4	53	9 US-09-864-761-43384	Sequence 398, App
45	38	38.4	61	11 US-09-764-872-398	

## ALIGNMENTS

## RESULT 1

US-03-674-913A-1  
; Sequence 1, Application US/09674913A  
; Publication No. US20030171266A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20030171266A1sk Hydro ASA  
; APPLICANT: Gaudernack, Gustav  
; APPLICANT: Eriksson, Jon Amund  
; APPLICANT: Moller, Mona  
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ub  
; TITLE OF INVENTION: Their Use  
; FILE REFERENCE: 001702.401600  
; CURRENT APPLICATION NUMBER: US/09/674,913A  
; CURRENT FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: PCT/NO99/00141  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: NO 19982098  
; PRIOR FILING DATE: 1998-05-08  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 1  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-674-913A-1

Query Match 100.0%; Score 99; DB 12; Length 19;  
Best Local Similarity 100.0%; Pred. No. 3.3e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NVPCHRMGRGRTSSKELA 19

Db 1 NVPCHRMGRGRTSSKELA 19

## RESULT 2

US-09-674-913A-2  
; Sequence 2, Application US/09674913A

```
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT FILING DATE: 2000-11-08
; PCT APPLICATION NUMBER: US/09/674.913A
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-3

Query Match          52.5%; Score 52; DB 12; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.099;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGHERMG 9
Db 9 NVPGHERMG 17
|||||||

RESULT 5
US-09-864-761-39237
; Sequence 39237, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Pean, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL IN
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Agomix-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/235,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT FILING DATE: 2000-11-08
; PCT APPLICATION NUMBER: US/09/674.913A
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-2

Query Match          100.0%; Score 99; DB 12; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.7e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGHERMGRTSSKELA 19
Db 9 NVPGHERMGRTSSKELA 27
|||||||

RESULT 3
US-09-674-913A-4
; Sequence 4, Application US/09674913A
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; APPLICANT: Moller, Mona
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT APPLICATION NUMBER: US/09/674.913A
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-4

Query Match          58.6%; Score 58; DB 12; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0075;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 MGRGTSSKELA 19
Db 1 MGRGTSSKELA 12
|||||||

RESULT 4
US-09-674-913A-3
; Sequence 3, Application US/09674913A
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT APPLICATION NUMBER: US/09/674.913A
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-3
```

```

1  PRIOR APPLICATION NUMBER: US 60/234,687
2  PRIOR FILING DATE: 2000-09-21
3  PRIOR APPLICATION NUMBER: US 09/608,408
4  PRIOR FILING DATE: 2000-06-30
5  PRIOR APPLICATION NUMBER: US 09/774,203
6  PRIOR FILING DATE: 2001-01-29
7  NUMBER OF SEQ ID NOS: 49117
8  SOFTWARE: Annonmax Sequence Listing Engine
9  SEQ ID NO 39237

```

Query Match 45.5%; Score 45; DB 9; Length 31;  
Best Local Similarity 50.0%; Pred. No. 2.5;  
Matches 9; Conservative 2; Mismatches 7; Indels

```

RESULT 6
US-09-764-891-4632
: Sequence 4632, Application US/09764891
: Publication No. US20030077808A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC006
: CURRENT APPLICATION NUMBER: US/09/764.891
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 10231
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4632
: LENGTH: 70
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (65)
: OTHER INFORMATION: Xaa equals any of the naturally occurring
: US-09-764-891-4632

```

RESULT 7  
US-10-291-851-70  
: Sequence 70, Application US/10291851  
: Publication No. US2003013838A1  
: GENERAL INFORMATION:  
: APPLICANT: Louie, Gordon V.  
: APPLICANT: Gajiwala, Ketan S.  
: APPLICANT: Buchanan, Sean Grant

```

: TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF MEMBERS OF
:
: TITLE OF INVENTION: THE P. coli COMA AND YGDB PROTEIN FAMILIES (COMA)
:
: FILE REFERENCE: 524982000700
:
: CURRENT APPLICATION NUMBER: US/10/291,851
:
: CURRENT FILING DATE: 2002-11-08
:
: PRIOR APPLICATION NUMBER: 60/337,683
:
: PRIOR FILING DATE: 2001-11-09
:
: NUMBER OF SEQ ID NOS: 96
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 70
:
: LENGTH: 124
:
: TYPE: PRT
:
: ORGANISM: P.aeruginosa
:
: US-10-291-851-70

```

```

RESULT 3
CS-10-156-761-11800
: Sequence 11800, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-2622
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 11800
: LENGTH: 480
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-11800

```

RESULT 9  
US-10-005-C57A-2  
: Sequence 2: Application US/10005057A  
: Publication No. US20020170087A1  
: GENERAL INFORMATION:  
: APPLICANT: Iao, Yumin  
: APPLICANT: Gordon-Kamm, William J.  
: APPLICANT: Shee, Ro  
: APPLICANT: Lowe, Keith S.  
: APPLICANT: Danilovskaya, Olga  
: APPLICANT: Mahajan, Pramod  
: APPLICANT: Kafalski, Antoni J.  
: APPLICANT: Sakai, Hajime

```

Query Match:      42.4%: Score 42; DB 10; Length 294;
  Nest Local Similarity 60.0%: Pred. No. 80;
Matches 9; Conservative 0; Mismatches 6; Indels

          QY      4 GHERMGRGRTSSKEL 18
              || | | | | |
          DD      184 GHGRGRGRSRSRKPL 198

RESULT 1:
US-10-092-263-8
; Sequence 8, Application US/10092263
; Publication No. US20030068811A1
; GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company

```

```

: TITLE OF INVENTION: SERORESELECTIVE REDUCTION OF SUBSTITUTED ACETOPHENON
:
: FILE REFERENCE: CT-2657NP
:
: CURRENT APPLICATION NUMBER: US/10/092,263
:
: CURRENT FILING DATE: 2002-06-04
:
: NUMBER OF SEQ ID NOS: 8
:
: SOFTWARE: PatentIn version 3.0
:
: SEQ ID NO 8
:
: LENGTH: 353
:
: TYPE: PRT
:
: ORGANISM: Pichia methanoclica
:
: US-10-092-263-8

```

```

US-10-092-263-8
  ORGANISM: Pichia methanocella

Query Match      42.4%   Score 42:   DB 15:   Length 353:
Best Local Similarity 44.4%   Pred. No. 96:
Matches 8; Conservative 2; Mismatches 8; Indels

CY      1 NVPGHERMGRTTSKEL 18
      ||||| | | | |
CB      1-5 NPPGHELVNRSGLSKHH 132

```

Wollman, Neil  
Thomson, Gerald H.  
Melton, Douglas A.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/945,182  
FILING DATE: 31-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/808,324  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazar, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5202-D  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617 498-8260  
TELEFAX: 617 876-5851  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:

US-09-945-182-34  
Query Match 42.4%; Score 42; DB 10; Length 388;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18  
DB 278 GHGRRGRSRCRKPL 232

## RESULT 13

US-09-945-182-28  
Sequence 28, Application US/09945192  
Patent No. US20020160494A1  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
Rosen, Vicki A.  
Wozney, John  
Thomsen, Neil  
Thomsen, Gerald H.  
Melton, Douglas A.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/945,182  
FILING DATE: 31-Aug-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/808,324  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazar, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5202-D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 498-8260  
TELEFAX: 617 876-5851

INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 28:

US-09-945-182-28  
Query Match 42.4%; Score 42; DB 10; Length 411;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18

DB 301 GHGRRGRSRCRKPL 315  
RESULT 14  
US-10-141-541-2  
Sequence 2, Application US/10141541  
Publication No. US20030167495A1  
GENERAL INFORMATION:  
APPLICANT: Pulst, Stefan M.  
TITLE OF INVENTION: SCA2 Knockout Animal and Methods of Use  
FILE REFERENCE: P-CF 5244  
CURRENT APPLICATION NUMBER: US/10/141,541;  
PRIOR FILING DATE: 2002-05-07  
PRIOR APPLICATION NUMBER: US 60/289,231  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 2  
LENGTH: 418  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-141-541-2

Query Match 42.4%; Score 42; DB 12; Length 418;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18  
DB 54 GRPGLGRGRNSSKGL 78

## RESULT 15

US-10-188-246-12  
Sequence 12, Application US/10188246  
Publication No. US20030087274A1  
GENERAL INFORMATION:  
APPLICANT: Anderson, David W.,  
Baldog, Ferenc,  
Casman, Stacie,  
Edinger, Shlomit,  
Gerlach, Valerie,  
Gorman, Linda,  
Li, Li,  
Malyankar, Uriel,  
Patturajan, Meera,  
Peyman, John,  
Shenoy, Suresh,  
Shimkets, Richard,  
Vernet, Corine A. M.,  
Voss, Edward  
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND  
FILE REFERENCE: 21402-397B US  
CURRENT APPLICATION NUMBER: US/10/188,246  
CURRENT FILING DATE: 2002-07-02  
PRIOR APPLICATION NUMBER: 60/303046  
PRIOR FILING DATE: 2001-07-05  
PRIOR APPLICATION NUMBER: 60/304502  
PRIOR FILING DATE: 2001-07-11  
PRIOR APPLICATION NUMBER: 60/305011  
PRIOR FILING DATE: 2001-07-12  
PRIOR APPLICATION NUMBER: 60/306085  
PRIOR FILING DATE: 2001-07-17  
PRIOR APPLICATION NUMBER: 60/326981  
PRIOR FILING DATE: 2001-10-14  
PRIOR APPLICATION NUMBER: 60/360923  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: 60/363636  
PRIOR FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/373063  
PRIOR FILING DATE: 2002-04-16  
NUMBER OF SEQ ID NOS: 60

```
; SOFTWARE: Custom
; SEQ ID NO 12
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-246-12

Query Match      42.4% Score 42; DB 15; Length 450;
Best Local Similarity 60.0% Pred. No. 1.2e+02;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      4 GHERMGRGRTSSKEL 18
      || || || || ||
Db      340 GHRGRGRSRCSRKPL 354

Search completed: October 1, 2003, 09:50:21
Job time : 65 secs
```

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:35:19 ; Search time 97 Seconds  
 (without alignments)  
 50.546 Million cell updates/sec

Title: US-09-674-913A-1  
 Perfect score: 99  
 Sequence: 1 NYPGHERMGRTSSKELA 19

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 330525

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

## Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protein:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.5	51.0	251	5 Q9V166	Q9V166 drosophila
2	50	50.5	72	16 Q9CIP3	Q9CIP3 lactococcus
3	46	46.5	342	16 Q97DL5	Q97DL5 clostridium
4	46	46.5	441	4 Q50149	Q50149 homo sapien
5	45.5	46.0	533	16 Q8CN90	Q8CN90 staphylococ
6	45	45.5	157	16 Q9HTU8	Q9HTU8 pseudomonas
7	45	45.5	331	10 Q9LH70	Q9LH70 arabidopsis
8	44.5	44.9	484	16 Q8XMY9	Q8XMY9 clostridium
9	44	44.4	2533	5 Q8IDH3	Q8IDH3 plasmodium
10	43.5	43.9	498	16 Q54116	Q54116 streptomyces
11	43	43.4	70	16 Q8E663	Q8E663 streptococ
12	43	43.4	70	16 Q8E018	Q8E018 streptococ
13	43	43.4	79	16 Q97RM2	Q97RM2 streptococ
14	43	43.4	79	16 Q8D0G2	Q8D0G2 streptococ
15	43	43.4	154	16 Q8PNB6	Q8PNB6 xanthomonas
16	43	43.4	244	4 Q8TF63	Q8TF63 homo sapien

17	43	43.4	424	3 Q74328	Q74328 schizosacch
18	43	43.4	441	11 Q99M11	Q99M11 mus musculu
19	42.5	42.9	491	16 Q97RA9	Q97RA9 streptococ
20	42.5	42.9	491	16 Q8DQ73	Q8DQ73 streptococ
21	42.5	42.9	1200	11 Q8CBV3	Q8CBV3 mus musculu
22	42	42.4	37	4 Q8XV6	Q8XV6 homo sapien
23	42	42.4	113	6 Q9NLS5	Q9NLS5 capreolus c
24	42	42.4	242	16 Q98608	Q98608 rhizobium l
25	42	42.4	294	6 Q9BDW9	Q9BDW9 macaca fasc
26	42	42.4	325	3 Q43042	Q43042 schizosacch
27	42	42.4	349	11 Q8R252	Q8R252 mus musculu
28	42	42.4	349	11 Q9D5G4	Q9D5G4 mus musculu
29	42	42.4	349	11 Q8BZ83	Q8BZ83 mus musculu
30	42	42.4	419	11 Q97421	Q97421 mus musculu
31	42	42.4	427	4 Q9ONY7	Q9ONY7 homo sapien
32	42	42.4	438	13 Q8UUP2	Q8UUP2 xenopus lae
33	42	42.4	441	11 Q921B6	Q921B6 mus musculu
34	42	42.4	447	6 Q9BDW8	Q9BDW8 cercopithe
35	42	42.4	481	17 Q97BP5	Q97BP5 thermoplas
36	42	42.4	517	4 Q75530	Q75530 homo sapien
37	42	42.4	535	11 Q97462	Q97462 mus musculu
38	42	42.4	586	16 Q9A8S0	Q9A8S0 caulobacter
39	42	42.4	821	6 Q19060	Q19060 sagittinus oe
40	42	42.4	961	16 Q8YQV0	Q8YQV0 anabaena sp
41	42	42.4	962	16 Q9JYD2	Q9JYD2 neisseria m
42	42	42.4	962	16 Q9JTB5	Q9JTB5 neisseria m
43	42	42.4	1084	10 Q8SB87	Q8SB87 oryza sativ
44	42	42.4	1285	11 Q70305	Q70305 mus musculu
45	41.5	41.9	236	16 Q9KZM4	Q9KZM4 streptomyce

## ALIGNMENTS

RESULT: 1

Q9V166 PRELIMINARY: PRT: 251 AA.

AC Q9V166: Q95SF6:

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DI 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DI 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE CG31248 protein (GH28633p).

GN CG31248 OR CG2640.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Ephydroidea; Ephydroidea; Ephydroidea;

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaralides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chex L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champ M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter F.G., Helt J., Nelson C.R., Miklos G.L.G.,

Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balowin D.,

Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Reeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

Cherry J.M., Cawley S., Dahlke C., Davenport L.A., Davies P.,

de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

Codson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

Glodok A., Gong F., Gorrelli J.H., Gu Z., Guan P., Harris M.,

Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,

Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,



```
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Moutkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith L.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [12]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup I.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ithegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.:
RA "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Cambell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.E., Smith C.D.,
RA Tupy J.L., Bergman B., Bertram B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smitniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.:
RA "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [14]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.:
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [15]
RP SEQUENCE FROM N.A.
RA FlyBase:
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [16]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkeley:
RA Stapleton M., Brokstein P., Hong J., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.:
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL: AE003672; AAF54059.2; -
DR ENBL: AY060818; AAL28366.1; -
DR FlyBase: FBgn0051248; CG31248.
DR SEQUENCE 251 AA; 28830 MW; C7CFF58D4E079F9 CRC64;
Query Match 51.0%; Score 50.5; DB 5; Length 251;
Best Local Similarity 48.0%; Pred. No. 2.7;
Matches 1; Conservative 3; Mismatches 3; Indels 7; Gaps 1;
Oy 2 VPGHERMGRGRTSS-----KELA 19
:| |||:| ||:|
3 PGERMGRGRTSSKE 18
||:| ||| |||:|
Query Match 46.5%; Score 46; DB 16; Length 342;
Best Local Similarity 56.2%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Oy 3 PGERMGRGRTSSKE 18
||:| ||| |||:|
153 :PSHERGLGSLNSQTLTLTKELA 177
RESULT 2
OQCIP3 PRELIMINARY; PRT; 72 AA.
AC OQCIP3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein ydbC.
GN YDB_C OX L:0313
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RA "The complete genome sequence of the lactic acid bacterium Lactococcus
RA lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR ENBL: AE006268; AAK04411.1; -
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 72 AA; 8405 MW; C5041FC4543B8410 CRC64;
Query Match 50.5%; Score 50; DB 16; Length 72;
Best Local Similarity 60.0%; Pred. No. 0.88;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Oy 3 PGERMGRGRTSSKE 17
||:| ||| |||:|
46 PDREKMGKGITLSEE 60
||:| ||| |||:|
RES:LT 3
O97DL5 PRELIMINARY; PRT; 342 AA.
AC O97DL5;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Homolog of cell division GTPase FtsZ, diverged.
GN CAC3459.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Daly M.J.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm S., Smith D.R.;
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RA bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR ENBL: AE007843; AAK81388.1; -
DR InterPro: IPR000217; Tubulin.
DR PROSITE: PS00227; TUBULIN; 1.
KW Cell division; Complete proteome.
SQ SEQUENCE 342 AA; 37499 MW; D2288911A0485F00 CRC64;
Query Match 46.5%; Score 46; DB 16; Length 342;
Best Local Similarity 56.2%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Oy 3 PGERMGRGRTSSKE 18
||:| ||| |||:|
```

```

Db      57 PGTDGSGDRDRKSKEM 72

RESULT 4
ID00149
AC      O00149;
DT      01-JUL-1997 (TReMBLrel. 04, Created)
DT      01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Embryonic ECTODERM development protein homolog.
GN      EED.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RN      SEQUENCE FROM N.A.
RA      Peytavi R., Hong S.S., Gay B., Dupuy D'Angeac A., Seliq L., Benichou S.,
RA      Benichou S., Benarous K., Boulanger P.;
RT      "HEED, the product of the human homolog of the murine eed gene, binds
RT      to the matrix protein of HIV-1.";
RI      J. Biol. Chem. 274:1635-1645(1999).
DR      EMBL: U90651; AAD08714.1;
DR      EMBL: AF099032; AAD08815.1;
DR      InterPro: IPR000408; Reg_chnr_Condens.
DR      InterPro: IPR001680; WD40.
DR      Pfam: PF00400; WD40; 5.
DR      SMART: SM00320; WD40; 3.
DR      PROSITE: PS00626; RC1.2; 1.
DR      PROSITE: PS00678; WD_REPEATS.1; 1.
DR      PROSITE: PS00882; WD_REPEATS.2; 2.
DR      PROSITE: PS02094; WD_REPEATS_REGION.1; 1.
KW      Repeat; WD repeat.
SQ      SEQUENCE 441 AA; 50155 MW; 2EAEA5BEEEA56B0 CRC64;

Query Match 46.5%; Score 46; DB 4; Length 441;
Best Local Similarity 47.1%; Pred. No. 27;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy      1 NVPGHERMGRGRTSSKE 17
      | | | | | | | | | |
Db      61 NAPGRKSGKGGKWSKK 77

RESULT 5
QBCN90
ID      Q8CN90
AC      Q8CN90;
DT      01-MAR-2003 (TReMBLrel. 23, Created)
DT      01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      L-lactate permease lctP-like protein.
GN      SE1945.
OS      Staphylococcus epidermidis.
OC      Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX      NCBI_TaxID=1292;
[1]
RN      SEQUENCE FROM N.A.
RA      Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA      Chen Z., Wen Y.;
RI      Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RL      EMBL: AE016750; AAO05586.1;

Query Match 45.5%; Score 45; DB 16; Length 157;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

Qy      2 VPGHERM-----GRGRT 13
      | | | | | | | | | |
Db      85 LPGHERMARLSKLTIDLRVDYLRPGRGRT 118

RESULT 6
Q9HTU8
ID      Q9HTU8
AC      Q9HTU8;
DT      01-MAR-2001 (TReMBLrel. 16, Created)
DT      01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Hypothetical protein PA5246.
GN      PA5246.
OS      Pseudomonas aeruginosa.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC      Pseudomonadaceae; Pseudomonas.
OX      NCBI_TaxID=287;
[1]
RN      SEQUENCE FROM N.A.
RA      STRAIN-ATCC 15692 / PA01;
RA      MEDLINE=2043737; PubMed=10984043;
RA      Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner F.,
RA      Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik S.J., Lagrou M.,
RA      Carber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA      Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA      Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA      Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RI      "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RI      opportunistic pathogen.";
RL      Nature 406:959-964(2000).
DR      EMBL: AE004937; AAG08631.1;
DR      InterPro: IPR003736; DUF157.
DR      InterPro: IPR006683; Thioestr_supf.
DR      Pfam: PF03061; 4HBT; 1.
DR      TIGRFAMS: TIGR00369; unchar_dom_1; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 157 AA; 17005 MW; 1002F16B858253A3 CRC64;

Query Match 45.5%; Score 45; DB 16; Length 157;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

Qy      2 VPGHERM-----GRGRT 13
      | | | | | | | | | |
Db      85 LPGHERMARLSKLTIDLRVDYLRPGRGRT 118

RESULT 7
Q9LH70
ID      Q9LH70
AC      Q9LH70;
DT      01-OCT-2000 (TReMBLrel. 15, Created)
DT      01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT      01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE      GB|AAD2396.1.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
[1]
RN      SEQUENCE FROM N.A.
RA      STRAIN-Columbia;
RA      Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RI      Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=0307853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by nine-y Pt.
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002061; BAB02651.1; -.
DR InterPro: IPR004253; DUF231.
DR Pfam: PF03005; DUF231.1
SQ SEQUENCE 331 AA; 37697 MW; 165BE4B9EF7EB9B8 CRC64;

Query Match 45.5%; Score 45; DB 10; Length 331;
Best Local Similarity 56.2%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PCHRMGRGRTSSKEL 18
DB 253 PSHYKSSVGRGTSGREL 268

RESULT 8
OBXMY9
ID OBXMY9 PRELIMINARY; PRT: 484 AA.
AC OBXMY9
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Lysine decarboxylase.
GN DCLY OR CPE0549.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
ON NCBI_TaxID=1502;
RX STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003187; BAB0255.1; -.
DR InterPro: IPR000310; Decarboxylase.
DR Pfam: PF01276; OKR_DC_1; 1.
DR Pfam: PF03711; OKR_DC_1-C; 1.
DR PROSITE: PS00703; OKR_DC_1; 1.
KW Complete proteome.
SQ SEQUENCE 484 AA; 53971 MW; E33933E34C92EFAC CRC54;

Query Match 44.9%; Score 44.5; DB 16; Length 484;
Best Local Similarity 52.9%; Pred. No. 54;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 1 NYPGHERMGRGRTSSKE 17
DB 27 DVPGH-KQGRGNTELRE 42

RESULT 9
OBIDH3 PRELIMINARY; PRT: 2533 AA.
AC OBIDH3
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein, conserved.
GN PF13_0273.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Lise A., Barton A., Corton C.,
RA Herriman M., Paig A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Crimond D., Sanders M., Hayes R., Hall S., Quail M., Harrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AI844509; CAD52648.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2533 AA; 304302 MW; 42BB01D2645D26BD CRC64;

Query Match 44.4%; Score 44; DB 5; Length 2533;
Best Local Similarity 43.8%; Pred. No. 3.7e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 NYPGHERMGRGRTSSK 16
DB 408 NIPNHDHKKKKRSSTK 423

RESULT 10
OS4116 PRELIMINARY; PRT: 498 AA.
AC OS4116
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE SC10A5.25c protein.
GN SC05920 OR SC10A5.25c.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RI Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=9700351; PubMed=8843436;
RA Kedenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RI "Complete genome sequence of the model actinomycete Streptomyces
RI coelicolor A3(2).";
RJ Nature 417:141-147(2002).
DK EMBL: AL939125; CAAL6457.1; -.
DR HSSP: Q58083; 1HV8.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001254; Ser_protease_Try.

```

```

DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW ATP-binding; Helicase; Complete proteome.
SQ SEQUENCE 498 AA; 52678 MW; 79F02BC9D5C96D10 CRC64;

Query Match 43.9% Score 43.5; DB 16; Length 473;
Best local Similarity 55.0%; Pred. No. 81;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 0;

QY 1 NVPGERHMRGRCT-SSXELA 19
   : : : : : : : : : : : :
DB 103 SLAGROINGRGTCSGKTLA 122

RESULT 11
Q8E663 Q8E663 PRELIMINARY; PRT; 70 AA.
AC Q8E663;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
DE GBS0764.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEM316 / Serotype I;I;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusnik C., Buchrieser C., Chevalier F., Prangeul L.,
RA Maadec T., Zouline M., Couve E., Mallou L., Poyart C., Frieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766846; CAD46408.1; -
DR Sgallist; gbs0764; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 8106 MW; 693AF5432945686F CRC64;

Query Match 43.4% Score 43; DB 16; Length 70;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PGHERMGRGRTSSKE 17
   : : : : : : : : : :
DB 45 PDHTKMGKGIILSNE 59

RESULT 12
Q8E018 Q8E018 PRELIMINARY; PRT; 70 AA.
AC Q8E018;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
DE SAG0743.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2603 V/R / Serotype V;
RX MEDLINE=22222986; PubMed=12200547;
RA Tettelin H., Masiagnani V., Cieslewicz M.J., Eiser J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,

```

DE Conserved hypothetical protein.  
 GN SP0690.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria: Firmicutes: Lactobacillales: Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczyk L.C., Burgett S.,  
 DeHoff B.S., Estrom S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 Gilmore R., Glass J.S., Khoja H., Kraft A.R., Laquer R.E.,  
 LeBlanc D.J., Lee L.N., Letkowitz E.J., Lu J., Matsushima P.,  
 McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas I.I.,  
 Norris F.H., O'Gata M., Peery R.B., Robertson G.T., Rockey P.,  
 Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
 Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL: AE008446; AAK99494.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA: 9136 MW: 7CEAE89579461448 CRC64;

Query Match 43.4%; Score 43; DB 16; Length 79;  
 Best Local Similarity 53.3%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PGHERMGRTSSKE 17  
 DB 53 PDHTRMGKGTLSNE 67

## RESULT 15

Q8PNB6 PRELIMINARY; PRT: 154 AA.  
 AC Q8PNB6;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Hypothetical protein XAC1157.  
 GN XAC1157.  
 OS Xanthomonas axonopodis (pv. citri).  
 OC Bacteria: Proteobacteria: Gammaproteobacteria: Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,  
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.F.A.,  
 Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.F.,  
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.I.,  
 Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 Locali E.C., Machado M.A., Madelira A.M.B.N., Martincz-Rossi N.M.,  
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira F.C., Tezza R.I.D.,  
 Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 host specificities.";  
 RL Nature 417:459-463(2002).  
 DR EMBL: AE011745; AAM36029.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 154 AA: 17260 MW: FC8DDF515D2D7715 CRC64;

Query Match 43.4%; Score 43; DB 16; Length 154;

Best Local Similarity 61.5%; Pred. No. 29;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 PGHERMGRTSS 15  
 DB 7 FGHRLRRGRSS 19

Search completed: October 1, 2003, 09:40:41  
 Job time : 102 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: October 1, 2003, 09:34:45 : Search time 22 seconds  
(without alignments)  
40.614 Million cell updates/sec

Title: US-09-674-913A-1

Perfect score: 99  
Sequence: 1 NPGHERMGRGRTSSKELA 19

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	44.5	44.9	1 RCA_ANASP	P58555 anabaena sp
2	43	43.4	1 GDF7_MOUSE	P43029 mus musculus
3	42.5	42.9	1 RCA_ANASC	Q06721 anabaena sp
4	42	42.4	1 RECA_RHOCA	P42447 rhodobacter
5	42	42.4	1 RECA_PARDE	P95469 paracoccus
6	41	41.4	1 RS13_MYCHO	P45813 mycobacteri
7	41	41.4	1 RECA_ACHLA	P29225 acholoplasm
8	41	41.4	1 RECA_BACAN	Q9ap22 bacillus an
9	40	40.4	1 IGF2_BOVIN	P07456 bos taurus
10	40	40.4	1 IGF2_SHEEP	P10764 ovis aries
11	40	40.4	1 RECA_RHOSH	P32725 rhodobacter
12	40	40.4	1 CBIA_SALT_Y	P29946 salmonella
13	40	40.4	1 REB1_SCHPO	Q9p6h9 schizosacch
14	40	40.4	1 ROM_HUMAN	P52272 homo sapien
15	39.5	39.9	1 POLX_TOBAC	P10978 nicotiana t
16	39	39.4	1 CAS1_PIG	P39035 sus scrofa
17	39	39.4	1 RNC_CORGL	Q8nnv6 corynebacte
18	39	39.4	1 CYSM_BACSU	Q34476 bacillus su
19	39	39.4	1 MID_PTCT	P42754 petroselinu
20	39	39.4	1 RECA_PORGI	P77925 porphyromon
21	39	39.4	1 RECA_CHLTE	O52393 chlorobium
22	39	39.4	1 YAHK_ECOLI	P75691 escherichia
23	39	39.4	1 D12_GREAL	O81931 crepis aipi
24	39	39.4	1 PRIM_PSEAE	Q915w0 pseudomonas
25	39	39.4	1 XLNR_ASPNG	C42804 aspergillus
26	39	39.4	1 RPOD_SINAL	Q9thv5 sinapis alb
27	39	39.4	1 CCAS_RAT	O54898 rattus norv
28	38	38.4	1 Y4WH_RH1SN	P55686 rhizobium s
29	38	38.4	1 IGF2_PIG	P23695 sus scrofa
30	38	38.4	1 RS4_PSEAE	O52759 pseudomonas
31	38	38.4	1 C9U7_HUMAN	Q9btj7 homo sapien
32	38	38.4	1 FLAGG_BORBU	O51715 borrelia bu
33	38	38.4	1 PANB_CORGL	Q9x712 corynebacte

34 38 38.4 325 1 ADD\_AGRKT5 Q8uj05 agrobacteri  
35 38 36.4 335 1 CGD1\_ARATH P42751 arabidopsis  
36 38 38.4 347 1 RECA\_BACSU P16971 bacillus su  
37 38 38.4 348 1 RECA\_AOUPY P33542 aquifex pyr  
38 38 38.4 349 1 RECA\_BACHD Q9kaa7 bacillus ha  
39 38 38.4 351 1 RECA\_CHLAU C52384 chloroflexu  
40 38 38.4 360 1 MTDH\_ARATH P42734 arabidopsis  
41 38 38.4 375 1 FDH\_NEUCR Q07103 pyrospora  
42 38 38.4 406 1 AROA\_PYRFU Q8u3a0 pyrococcus  
43 38 38.4 413 1 RCA\_CUCSA Q01587 cucumis sat  
44 38 38.4 442 1 GLXD\_RHIME O87392 rhizobium m  
45 38 38.4 442 1 TBB\_EUGGR P12457 euglena gra

## ALIGNMENTS

RESULT 1  
PCA\_ANASP STANDARD; PRT; 414 AA.  
AC P58555:  
IT 28-FEB-2003 (Rel. 41, Created)  
LT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Ribulose biphosphate carboxylase/oxygenase activase (RuBisCO  
activase) (P8a).  
GN RCA\_OA ALR1533.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
PP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
Kawanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
Yasuda M., Tabata S.;  
RT \*Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120.\*;  
RL DNA Res. 8:205-213(2001).  
CC -1- FUNCTION: ACTIVATION OF RUBISCO (RUBISCO-1,5-BISPHOSPHATE  
CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT  
CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A  
CARBAMATE STRUCTURE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercia  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC .....  
CC EMBL; AP003586; BAB77859.1; -  
DR PIR; AG1997; AG1997.  
DR InterPro; IPR003959; AAA\_Attrase\_centri.  
DR Pfam; PF00004; AAA; 1.  
KW ATP-binding; Complete proteome.  
FT NE\_BIND 37 44 ATP (POTENTIAL).  
SQ SEQUENCE 414 AA: 46668 MW: E8359F10D8C558DA CRC64:  
44.9% Score 44.5; DB 1; Length 414;  
Best Local Similarity 39.3%; Pred. No. 5.3;  
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;  
Qy 1 NVFG-----HERMGRGRTSSKELA 19  
|:| | | | | | | | | |  
Db 23 NPGVRVPLILGTHGRGCKTFOCELA 50

RESULT 2

```

GDF7_MOUSE
ID GDF7_MOUSE STANDARD: PRT: 151 AA.
AC P43029;
DT 01-NOV-1995 (Rel. 32, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Growth/differentiation factor 7 precursor (GDF-7) (Fragment).
GN GDF7 OR GDF-7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=94195427; PubMed=8145850;
RA Storm E.E., Huynh T.V., Copeland N.G., Jenkins N.A., Kingsley D.M.,
RA Lee S.-J.;
RT *Limb alterations in brachypodism mice due to mutations in a new
RT member of the TGF beta-superfamily.*;
RL Nature 368:639-643(1994)
CC -!- SUBUNIT: Homodimer; disulfide-linked (By similarity).
CC -!- SIMILARITY: Belongs to the TGF-beta family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U08339; AAA18780.1;
CC PIR: S43296; S43296.
CC HSP: P12643; 38MP.
CC GDI: GDI:95690; GdiF.
CC InterPro: IPR002400; GF_cysknot.
CC InterPro: IPR001839; TGFb.
CC Pfam: PFC0019; TGF-beta; 1.
CC PRINTS: PR00438; GFCYSKNOT.
CC ProDom: PD000357; TGPB; 1.
CC SMART: SM00204; TGFb; 1.
CC PROSITE: PS00250; TGF_BETA_1; 1.
CC Growth factor; Cytokine; Glycoprotein.
FT NON_TER 1 1
FT PROPEP <1 5 POTENTIAL.
FT CHAIN 6 151 GROWTH/DIFFERENTIATION FACTOR 7.
FT DISULFID 50 116 BY SIMILARITY.
FT DISULFID 79 148 BY SIMILARITY.
FT DISULFID 83 150 BY SIMILARITY.
FT DISULFID 115 115 INTERCHAIN (BY SIMILARITY).
FT DOMAIN 1 5 POLY-ARG.
FT DOMAIN 16 41 POLY-GLY.
SQ SEQUENCE 151 AA; 15697 MW; 0E496ACB5827759 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 151;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 6; Indels 2; Gaps 0;

QY 4 GHGRGRGRTSSKEL 18
DB 41 GHGRGRGRTSSKSL 55

RESULT 3
RCA_ANASC
ID RCA_ANASC STANDARD: PRT: 415 AA.
AC Q06721;
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribulose biphosphate carboxylase/oxygenase activase (RuBisCo
DE activase) (RA).

```

```

RCA_
OS Anabaena sp. (strain CA / ATCC 33047).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=52271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93222475; PubMed=8467074;
RA L.L., Gibson J.L., Tabita F.R.;
RT "The Rubisco activase (rca) gene is located downstream from rbcS in
RT Anabaena sp. strain CA and is detected in other Anabaena/Nostoc
RT strains.*";
RL Plant Mol. Biol. 21:753-764(1993).
CC -!- FUNCTION: ACTIVATION OF RUBISCO (RUBULOSE-1,5-BISPHOSPHATE
CC CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT
CC CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A
CC CARBAMATE STRUCTURE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X67942; CAA48129.1;
CC InterPro: IPR003959; AAA_Atpase_centri.
CC Pfam: PF00004; AAA; 1.
CC ATP-binding. 37 44
FT NP_BINS 37 44 ATP (POTENTIAL).
SQ SEQUENCE 415 AA; 46594 MW; 1CF71296D94A892D CRC64;

Query Match 42.9%; Score 42.5; DB 1; Length 415;
Best Local Similarity 39.3%; Pred. No. 12;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 NVPS-----HERMGRGRTSSKELA 19
DB 23 NLPGVRVPLIIGIRGKGKGFQCEIA 50

RESULT 4
RCA_RHOCA
ID RCA_RHOCA STANDARD: PRT: 355 AA.
AC P4247;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rca protein (Recombinase A).
GN RCA.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=J50;
RX MEDLINE=95331467; PubMed=7607398;
RA Fernandez de Henestrosa A.R., Rivera E., Barbe J.;
RT "Non-reciprocal regulation of Rhodobacter capsulatus and Rhodobacter
RT sphaeroides rca genes expression.*";
RL FEMS Microbiol. Lett. 129:175-181(1995).
CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, THE ATP-DEPENDENT HYBRIDIZATION OF
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEVA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the rca family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X82183; CAA57673.1; --  
 DR PIR: S49464; S49464.  
 DR HSSP: P03017; 2REB.  
 DR HAMAP: MF\_00268; --; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001553; RecA.  
 DR Pfam: PF00154; RecA\_1.  
 DR PRINTS: PR00142; RECA.  
 DR SMART: SM00382; AAA; 1.  
 DR PRODOM: PD000229; RecA; 1.  
 DR PROSITE: PS00321; RECA\_1; 1.  
 DR PROSITE: PS00162; RECA\_2; 1.  
 DR PROSITE: PS0163; RECA\_3; 1.  
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.  
 FT NP\_BIND 78 85 ATP (BY SIMILARITY).  
 SQ SEQUENCE 355 AA; 38027 MW; 64E498A33A2DE77E CRC64;

Query Match 42.4%; Score 42; DB 1; Length 355;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKE 17  
 : |||: ||| :  
 DB 306 GDERIGQGRENAKO 319

## RESULT 5

RECA\_PARDE ID RECA\_PARDE STANDARD; PRT; 356 AA.  
 AC P95469;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE RECA protein (Recombinase A).  
 GN RECA.  
 OS Paracoccus denitrificans.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
 OC Rhodobacteraceae; Paracoccus.  
 OX NCBI\_TaxID=266;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=DSM 413;  
 RX MEDLINE=97211997; PubMed=9119195;  
 RA Fernandez de Henestrosa A.R., del Rey A., Tarrago R., Barbe J.;  
 RT Cloning and characterization of the recA of Paracoccus denitrificans  
 RT and construction of a recA-deficient mutant.;  
 RL FEMS Microbiol. Lett. 147:209-213(1997).

CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF  
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED  
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF  
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING  
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the recA family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U59631; AAB51561.1; --  
 DR HSSP: P03017; 2REB.  
 DR HAMAP: MF\_00268; --; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.

DR InterPro: IPR001553; RecA.  
 DR Pfam: PF00154; recA; 1.  
 DR PRINTS: PR00142; RECA.  
 DR PRODOM: PD000229; RECA; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00321; RECA\_1; 1.  
 DR PROSITE: PS00162; RECA\_2; 1.  
 DR PROSITE: PS0163; RECA\_3; 1.  
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.  
 FT NP\_BIND 78 85 ATP (BY SIMILARITY).  
 SQ SEQUENCE 356 AA; 38092 MW; FAC0C06A7424EBBC CRC64;

Query Match 42.4%; Score 42; DB 1; Length 356;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKE 17  
 : |||: ||| :  
 DB 306 GDERIGQGRENAKO 319

## RESULT 6

RS13\_MYCBO ID RS13\_MYCBO STANDARD; PRT; 123 AA.  
 AC P45813;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 30S ribosome; protein S13.  
 DE RPSM.  
 GN Mycobacterium bovis.  
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BCG;  
 RX MEDLINE=96200849; PubMed=8621083;  
 RA Dubnau R., Soares S., Huang T.J., Jacobs W.R. Jr.;  
 RT Overproduction of mycobacterial ribosomal protein S13 induces  
 RT catalase/peroxidase activity and hypersensitivity to isoniazid in  
 RT Mycobacterium smegmatis.\*;  
 RL Gene 170:17-22(1996).

CC -!- FUNCTION: INVOLVED IN THE BINDING OF FMET-TRNA AND, HENCE, IN THE  
 CC INITIATION OF TRANSLATION (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U15140; AAB17598.1; --  
 DR InterPro: IPR001892; Ribosomal\_S13.  
 DR Pfam: PF00416; Ribosomal\_S13; 1.  
 DR PRODOM: PD001363; Ribosomal\_S13; 1.  
 DR PROSITE: PS00646; RIBOSOMAL\_S13\_1; 1.  
 DR PROSITE: PS0159; RIBOSOMAL\_S13\_2; 1.  
 KW Ribosomal protein.

SQ SEQUENCE 123 AA; 14407 MW; 8FE284FC620EFF245 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 123;  
 Best Local Similarity 38.5%; Pred. No. 5.6;  
 Matches 10; Conservative 4; Mismatches 4; Indels 8; Gaps 1;

QY 1 NYPGHERM-----GKGRTSSEL 18  
 : |||: ||| :  
 DB 8 DLPRHRMEVALTYIFGIGRTSNEI 33



```

RESULT 7
RECA_ACHLA
ID RECA_ACHLA STANDARD: PRT: 331 AA.
P29225:
D* 01-DEC-1992 (Rel. 24, Created)
D* 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RECA protein (Recombinase A).
RECA.
OS Acholeplasma laidlawii.
OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;
OC Acholeplasmataceae; Acholeplasma.
OX NCBI_TaxID=2148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL, K2, and 8195;
RX MEDLINE=92121115; PubMed=1732213;
R* Cloning and DNA sequence of a mycoplasmal recA gene.";
R* J. Bacteriol. 174:778-784(1992).
RL
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC -!- SIMILARITY: Belongs to the recA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M81465; AAB59011.1; -.
CC PIR: A42602; A42602.
CC HSSP: P26345; IG18.
CC HAMAP: MF_00268; -.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR001553; RecA.
CC Pfam: PF00154; recA; 1.
CC PRINTS: PR00142; RECA.
CC PRODOM: PD000229; RecA; 1.
CC SMART: SM00382; AAA; 1.
CC PROSITE: PS00321; RECA_1; 1.
CC PROSITE: PS0162; RECA_2; 1.
CC PROSITE: PS0163; RECA_3; 1.
CC DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
CC NP_BIND 66 73
CC VARIANT 295 331 MISSING (IN STRAIN 8195).
CC SEQUENCE 331 AA; 35516 MW; 60F3EA59CED83FE2 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 331;
Best Local Similarity 42.9%; Pred. No. 16;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 3;

QY 4 GHERMGRGRTSSKE 17
| : : : : :
Db 294 GEKLGQGRDNQAK 307

RESULT 8
RECA_BACAN
ID RECA_BACAN STANDARD: PRT: 343 AA.
AC O9AP22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE RecA protein (Recombinase A).
RECA.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Stearns;
R* Ko M., Kim J.C., Park C.;
R* Cloning of recA gene from Bacillus anthracis.";
R* Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RL
CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC -!- SIMILARITY: Belongs to the recA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF229167; AAK00736.1; -.
CC HSSP: P26345; IG19.
CC HAMAP: MF_00268; -.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR001553; RecA.
CC Pfam: PF00154; recA; 1.
CC PRINTS: PR00142; RECA.
CC PRODOM: PD000229; RecA; 1.
CC SMART: SM00382; AAA; 1.
CC PROSITE: PS00321; RECA_1; 1.
CC PROSITE: PS0162; RECA_2; 1.
CC PROSITE: PS0163; RECA_3; 1.
CC DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
CC NP_BIND 64 71
CC VARIANT 295 331 ATP (POTENTIAL).
CC SEQUENCE 343 AA; 37296 MW; 06CE4CFE9C1BBB48 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 343;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 ERMGRGRTSSKE 17
| : : : : :
Db 294 ERLGQGRNSKQ 305

RESULT 9
IGF2_BOVIN
ID IGF2_BOVIN STANDARD: PRT: 155 AA.
AC P07456;
CT 0; APR-1988 (Rel. 07, Created)
ET 0; MAY-1992 (Rel. 22, Last sequence update)
ET 0; FEB-1996 (Rel. 33, Last annotation update)
DE Insulin-like growth factor II precursor (IGF-II) (Erythrotropin).
DE (fragment).
DE IGF2.
GN IGF2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 6-155 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90356421; PubMed=2388846;
RA Brown W.M., Dzrieglewska K.M., Foreman R.C., Saunders N.R.;
RT "The nucleotide and deduced amino acid sequences of insulin-like
RT growth factor II cDNAs from adult bovine and fetal sheep liver.";
RE Nucleic Acids Res. 18:4614-4614(1990).
RL
OS

```



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U00666; AAB60626.1; JOINED.  
 DR EMBL: U00666; AAB60626.1; JOINED.  
 DR EMBL: X15248; CAA33324.1; JOINED.  
 DR EMBL: X53554; CAA37621.1; JOINED.  
 DR EMBL: M89788; AAA31548.1; JOINED.  
 DR EMBL: M89789; AAA31549.1; JOINED.  
 DR EMBL: X55638; CAA39163.1; JOINED.  
 DR PIR: S04858; S04858.  
 DR HSSP: P01344; IGF2.  
 DR InterPro: IPR004825; Ins/IGF/relax.  
 DR Pfam: PF00049; Insulin; 1.  
 DR PRINTS: PRO0277; INSULIN.  
 DR SMART: SM00078; IIGF; 1.  
 DR PROSITE: PS00262; INSULIN; 1.  
 KW Insulin family; Mitogen; Growth factor; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 91 INSULIN-LIKE GROWTH FACTOR II.  
 FT DOMAIN 25 52 B.  
 FT DOMAIN 53 64 C.  
 FT DOMAIN 65 85 A.  
 FT DOMAIN 86 91 D.  
 FT PROPEP 92 179 E PEPTIDE.  
 FT DISULFID 33 71 BY SIMILARITY.  
 FT DISULFID 45 84 BY SIMILARITY.  
 FT DISULFID 70 75 BY SIMILARITY.  
 FT CONFLICT 46 47 GD -> DG (IN REF. 5).  
 SQ SEQUENCE 179 AA; 19616 MW; 7B369AE57F2E4378 CRC64;

Query Match 40.4%; Score 40; DB 1; Length 179;  
 Best Local Similarity 52.9%; Pred. No. 12;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 VQHEHMGRTSSKEL 18  
 : : : : :  
 DB 130 LPAFLRARRGTLAKEL 146

## RESULT 11

RECA\_RHOSH  
 ID RECA\_RHOSH STANDARD; PRT: 343 AA.  
 AC P32725;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE RecA protein (Recombinase A).  
 GN RECA.  
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
 OC Rhodobacteraceae; Rhodobacter.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=APCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
 RX MEDLINE=94104596; PubMed=8277942;  
 RA Calero S., Fernandez de Henestrosa A.R., Barbe J.,  
 RT "Molecular cloning, sequence and regulation of expression of the recA  
 RT gene of the phototrophic bacterium Rhodobacter sphaeroides.",  
 RL Mol. Gen. Genet. 242:116-120(1994).  
 CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF  
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED  
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF  
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING  
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the recA family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X72705; CAA51258.1; JOINED.  
 DR PIR: S41560; S41560.  
 DR HSSP: P03017; 2REB.  
 DR HAMAP: MF\_00268; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001553; RECA.  
 DR Pfam: PF00154; RECA; 1.  
 DR PRINTS: PH00142; RECA.  
 DR PRODOM: PD000229; RECA; 1.  
 DR SMART: SMG0382; AAA; 1.  
 DR PROSITE: PS00321; RECA\_1; 1.  
 DR PROSITE: PS0162; RECA\_2; 1.  
 DR PROSITE: PS0163; RECA\_3; 1.  
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.  
 FT NP\_BIND 64 71 ATP (BY SIMILARITY).  
 SQ SEQUENCE 343 AA; 36698 MW; 40584E781E5CAEDF CRC64;

Query Match 40.4%; Score 40; DB 1; Length 343;  
 Best Local Similarity 53.8%; Pred. No. 25;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 4 GHEHMGRTSSK 16  
 : : : : :  
 DB 292 GDERIGGRENK 304

## RESULT 12

CBIA\_SALTY  
 ID CBIA\_SALTY STANDARD; PRT: 459 AA.  
 AC P29946;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cobyric acid A.C-diamide synthase.  
 GN CBIA OR STM2035.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2;  
 RX MEDLINE=93273696; PubMed=8501034;  
 RA Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S.,  
 RA Church G.M.;  
 RT "Characterization of the cobalamin (vitamin B12) biosynthetic genes  
 RT of Salmonella typhimurium.",  
 RL J. Bacteriol. 175:3303-3316(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.",  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE OF 1-106 FROM N.A.  
 RC STRAIN=LT2;  
 RX MEDLINE=92244044; PubMed=1374146;  
 RA Richter-Dahlfors A.A., Andersson D.I.;  
 RT "Cobalamin (vitamin B12) repression of the Cob operon in Salmonella  
 RT typhimurium requires sequences within the leader and the first

```

RT translated open reading frame."
RL MOL. Microbiol. 6:743-749(1992).
CC -!- FUNCTION: RESPONSIBLE FOR THE AMIDATION OF CARBOXYLIC GROUPS AT
CC POSITION 1 AND C OF EITHER COBIRINIC ACID OR HYDROGENOBIRINIC ACID.
CC NH(2) GROUPS ARE PROVIDED BY GLUTAMINE, AND ONE MOLECULE OF ATP
CC IS HYDROGENOLYZED FOR EACH AMIDATION.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: Belongs to the cobB/cobQ family. CobB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: L12006; AAA27252.1; .
CC EMBL: AF008789; AAL20939.1; .
CC EMBL: X63012; CAN44740.1; .
CC PIR: S20553; S20553.
CC DR StyGene: SG10034; cbIA.
CC DR HAMAP: MF_00027; . 1.
CC DR InterPro: IPR004484; CbiA_P.
CC DR InterPro: IPR002586; CbiA_P.
CC DR Pfam: PF01656; CbiA; 1.
CC DR TIGRFAMs: TIGR00379; cobB; 1.
CC KW Cobalamin biosynthesis; Porphyrin biosynthesis; Complete protome.
CC FT CONFLICT 36 36 R > P (IN REF. 1).
CC FT CONFLICT 111 111 M > I (IN REF. 2).
CC FT CONFLICT 128 128 V > I (IN REF. 1).
CC FT CONFLICT 133 133 A > T (IN REF. 1).
CC SQ SEQUENCE 459 AA: 50036 MW: 51FE1B5A37F5C14 CRC64:
Query Match 40.4%; Score 40; DB 1; Length 459;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 VPGHERMGRGRT 13
:||||:|:|:|
DB 358 IPGHSKMKRLT 169
RESULT: 13
ID REB1 SCHPO STANDARD; PRT: 504 AA.
AC Q9P6H9; O94422;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-binding protein reb1.
GN REB1 OR SPBC1198.11C OR SPBC660.01C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_taxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE-97169306; PubMed-9016645;
RA Zhao A., Guo A., Liu Z., Pape L.;
RT "Molecular cloning and analysis of Schizosaccharomycetes pombe Reb1p:
RT sequence-specific recognition of two sites in the far upstream rDNA
RT intergenic spacer."
RL Nucleic Acids Res. 25:904-910(1997).
RN [2]
RP SEQUENCE FROM N.A.
RN STRAIN-972;
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourkos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabe, C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borycz K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl I.M.,
RA Kyrer P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu F., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ossery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe."
CC -!- FUNCTION: DNA-binding protein that recognizes sites within both
CC the enhancer and the promoter of rRNA transcription, as well as
CC upstream of many genes transcribed by RNA polymerase II. Has a
CC role in the termination of RNA polymerase I catalyzed
CC transcription.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 2 Myb-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U33010; . NOT ANNOTATED_CDS.
CC EMBL: AL355921; CAB01186.1; .
CC EMBL: AL034563; CAA2522.1; .
CC GeneDB-SPombe: SUBC1198.11C; .
CC InterPro: IPR001005; Myb_DNA_Binding.
CC Pfam: PF00249; myb_DNA-binding; 2.
CC SMART: SM00717; SANT; 2.
CC PROSITE: PS00037; MYB_1; 2.
CC PROSITE: PS00334; MYB_2; 1.
CC PROSITE: PS00900; MYB_3; 2.
CC KW Nuclear protein; DNA-binding; Transcription regulation; Repeat.
CC DNA_BIND 313 357 MYB 1.
CC FT DNA_BIND 365 418 MYB 2.
CC SQ SEQUENCE 504 AA: 58448 MW: 667C45F453E4F916 CRC64:
Query Match 40.4%; Score 40; DB 1; Length 504;
Best Local Similarity 37.5%; Pred. No. 39;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 NVPGHERMGRGRTSSK 16
:||||:|:|:|
DB 223 NIPGYEKYSRKNSSOR 238
RESULT: 14
ID ROM_HUMAN STANDARD; PRT: 730 AA.
AC P52272;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein M (hnRNP M).
GN HNRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93181232; PubMed=8441556;  
 RA Datar K.V., Dreyfuss G., Swanson M.S.;  
 RI "The human hnRNP M proteins: identification of a methionine/arginine-  
 rich repeat motif in ribonucleoproteins.";  
 RL Nucleic Acids Res. 21:439-446(1993).  
 RN [2]  
 RP SEQUENCE OF 1-59 FROM N.A.  
 RX MEDLINE=96292257; PubMed=8692593;  
 RA Gattoni R., Mahe C., Mahe P., Fischer N., Mattei M.G., Stevenin C.,  
 Fuchs J.-P.;  
 RI "The human hnRNP-M proteins: structure and relation with early heat  
 shock-induced splicing arrest and chromosome mapping.";  
 RL Nucleic Acids Res. 24:2535-2542(1996).  
 CC - FUNCTION: PRE-MRNA BINDING PROTEINS IN VIVO, AND THEY BIND ACTIVELY  
 TO POLY(G) AND POLY(U) RNA HOMOPOLYMERS IN VITRO. INVOLVED IN  
 SPLICING.  
 CC - SUBCELLULAR LOCATION: Nuclear.  
 CC - ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Experimental confirmation may be lacking for some  
 isoforms;  
 CC Name=M4;  
 CC IsoId=P52272-1; Sequence=Displayed;  
 CC Name=M1-M2;  
 CC IsoId=P52272-2; Sequence=VSP\_005845;  
 CC Name=M3;  
 CC IsoId=P52272-3; Sequence=Not described;  
 CC - SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: L03532; AAA36192.1;  
 DR PIR: S35532; S35532.  
 DR HSP: P1340; ICVJ.  
 DR Genbank: HGNC:5046; HNRPM.  
 DR GK: P52272;  
 DR MIM: 160994;  
 DR GO: GO:0005887; C: integral to plasma membrane; TAS.  
 DR GO: GO:0005624; C: membrane fraction; TAS.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam: PF00076; Rrm\_3.  
 DR SMART: SM00360; RRM; 3.  
 DR PROSITE: PS0102; RRM; 3.  
 DR PROSITE: PS0030; RRM\_RNP\_1; FALSE\_NEG.  
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;  
 KW Alternative splicing.  
 FT DOMAIN 71 149 RNA-BINDING (RRM) 1.  
 FT DOMAIN 204 281 RNA-BINDING (RRM) 2.  
 FT DOMAIN 390 396 POLY-GLY.  
 FT DOMAIN 400 608 27 X 6 AA REPEATS OF G-[IL]-[DE]-R-M-  
 [GA].  
 FT REPEAT 400 405 1.  
 FT REPEAT 407 412 2.  
 FT REPEAT 415 420 3.  
 FT REPEAT 426 431 4.  
 FT REPEAT 433 438 5.  
 FT REPEAT 440 445 6.  
 FT REPEAT 446 451 7.  
 FT REPEAT 453 458 8.  
 FT REPEAT 461 466 9.  
 FT REPEAT 468 473 10.  
 FT REPEAT 475 480 11.  
 FT REPEAT 482 487 12.

FT REPEAT 493 498 13.  
 FT REPEAT 500 505 14.  
 FT REPEAT 507 512 15.  
 FT REPEAT 514 519 16.  
 FT REPEAT 521 526 17.  
 FT REPEAT 528 533 18.  
 FT REPEAT 540 545 19.  
 FT REPEAT 547 552 20.  
 FT REPEAT 554 559 21.  
 FT REPEAT 562 567 22.  
 FT REPEAT 567 572 23.  
 FT REPEAT 575 580 24.  
 FT REPEAT 580 585 25.  
 FT REPEAT 588 593 26.  
 FT REPEAT 603 608 27.  
 FT DOMAIN 612 616 POLY-GLY.  
 FT DOMAIN 653 729 RNA-BINDING (RRM) 3.  
 FT VARSPLIC 160 198 Missing (in isoform M1-M2).  
 FT CONFLICT 24 34 /FTID=VSP\_005845.  
 FT SEQUENCE 730 AA; 77470 MW; H2236BE1D34AE27 CRC64;  
 Query Match 40.48; Score 40; DB 1; Length 730;  
 Best Local Similarity 72.78; Pred. No. 58;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 NVPGHERMGRG 11  
 DB 397 SVPGIERMGP 407  
 :||| ||||| I  
 RESULT 15  
 POLX IOBAC STANDARD; PRT; 1328 AA.  
 AC P10978;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DI 01-JUL-1989 (Rel. 11, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Retrovirus-related Pol polyprotein from transposon TNT 1-94 [Contains:  
 DE Protease (EC 3.4.23.-); Reverse transcriptase (EC 2.7.7.49);  
 DE Endonuclease].  
 OS Nicotiana tabacum (Common tobacco).  
 CC Eukaryota; Viridiplantae; Embryophyta; Streptophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.  
 CC NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89097311; PubMed=2536143;  
 RA Grandbastien M.-A., Spielmann A., Caboche M.;  
 RI "Entl, a mobile retroviral-like transposable element of tobacco  
 isolated by plant cell genetics.";  
 RL Nature 337:376-380(1989).  
 CC - CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate  
 + (DNA)(N).  
 CC - SIMILARITY: HIGH, WITH DROSOPHILA COP-A ELEMENT.  
 CC - SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.  
 CC - SIMILARITY: Contains 1 CCHC-type zinc finger.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: X13777; CAA32025.1;  
 DR PIR: S04273; S04273.  
 DR MEROPS: A11.002;  
 DR InterPro: IPR001995; Aspprotease\_rtrv.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR001878; Znf\_CCHC.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:35:49 ; Search time 39 Seconds  
(without alignments)  
46.851 Million cell updates/sec

Title: US-09-674-913A-1

Perfect score: 99  
Sequence: 1 NPGHERMGRGRTSSKELA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_76:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	50.5	72	2 A86664	hypothetical prote
2	46	46.5	342	2 A97325	homolog of cell di
3	45	45.5	157	2 H82989	conserved hypothet
4	44.5	44.9	414	2 A41997	ribulose-bisphosph
5	43.5	43.9	498	2 T34592	DEAD-box RNA helic
6	43	43.4	79	2 F95090	conserved hypothet
7	43	43.4	79	2 B97958	conserved hypothet
8	43	43.4	151	2 S43296	bone morphogenetic
9	43	43.4	244	2 JC7801	dendritic cell-ass
10	43	43.4	424	2 T39524	hypothetical prote
11	42.5	42.9	415	2 S33627	ribulose-bisphosph
12	42.5	42.9	491	2 G95105	lysine decarboxyla
13	42.5	42.9	491	2 H97973	lysine decarboxyla
14	42	42.4	326	2 T40353	probable ribosomal
15	42	42.4	355	2 S49464	recombination prot
16	42	42.4	586	2 B87408	tetracycline resist
17	42	42.4	961	2 AE2270	excinuclease ABC c
18	42	42.4	962	2 A81817	translation initia
19	42	42.4	962	2 C81060	translation initia
20	42	42.4	1285	2 T4171	ataxin-2 - mouse
21	41.5	41.9	355	2 AC2768	alcohol dehydrogen
22	41.5	41.9	368	2 E97548	alcohol dehydrogen
23	41	41.4	235	2 T36923	hypothetical prote
24	41	41.4	302	2 AB3573	oligopeptide trans
25	41	41.4	331	2 A42602	recombination prot
26	40	40.4	69	2 A31147	hypothetical prote
27	40	40.4	69	2 AD1506	hypothetical prote
28	40	40.4	155	1 IGB02	insulin-like growt
29	40	40.4	162	2 F96496	probable myosin he

30 40 40.4 179 2 S04858 insulin-like growt  
31 40 40.4 194 2 AB3605 transcription regu  
32 40 40.4 197 2 B72623 hypothetical prote  
33 40 40.4 215 2 T49743 probable rfm-type  
34 40 40.4 240 2 H90778 DNA-binding protei  
35 40 40.4 242 2 C85640 hypothetical prote  
36 40 40.4 256 2 D87149 conserved hypothet  
37 40 40.4 280 2 A48488 saccharide biosynt  
38 40 40.4 285 2 G83934 hypothetical prote  
39 40 40.4 342 2 AD2653 alcohol dehydrogen  
40 40 40.4 342 2 H97434 alcoh dehydrogen  
41 40 40.4 343 2 S41560 recombination prot  
42 40 40.4 423 2 A70947 hypothetical prote  
43 40 40.4 439 2 T02453 hypothetical prote  
44 40 40.4 447 2 T22163 hypothetical prote  
45 40 40.4 455 2 G70113 replicative DNA he

#### ALIGNMENTS

##### RESULT 1

A86664  
hypothetical protein ydbc [imported] - Lactococcus lactis subsp. lactis (strain IL  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: A86664  
R:Solotkin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmme, K.; Weissenbach, J.;  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lac  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: A86664  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-72 <STO>  
A:Cross-references: GB:AE005176; PID:g:2723:76; PIDN:AAK04411.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ydbc

Query Match 50.5%; Score 50; DB 2; Length 72;  
Best Local Similarity 50.0%; Pred. No. 0.35;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 PGRHMRGRITSSKE 17  
I I I I I I I I I  
DB 46 PDHMKMGK:TLSEE 60

##### RESULT 2

A97325  
homolog of cell division GTPase FtsZ, diverged [imported] - Clostridium acetobutyli  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: A97325  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteri  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: A97325  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-342 <NUP>  
A:Cross-references: GB:AE001437; PIDN:AAK01388.1; PID:g15026550; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3459

Query Match 46.5%; Score 46; DB 2; Length: 342;  
Best Local Similarity 56.2%; Pred. No. 7.6;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```
QY 3 PCHERMGRGRTSSKEL 18
      || : || || |||
Db 57 PGIDSGGRDR*KSREM 72

RESULT 3
H82989
conserved hypothetical protein PA5246 [imported] - Pseudomonas aeruginosa (strain: PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H82989
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; R-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H82989
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <STOS>
A:Cross-references: GB:AE004937; GB:AE004091; NID:g9951553; PIDN:AA8C8631.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA5246

Query Match 45.5%; Score 45; DB 2; Length 157;
Best Local Similarity 36.7%; Pred. No. 5.1;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

QY 2 VPGHERM-----GRGRT 13
      |||||
Db 89 LPGHERMARLSKGLIDLRVDYLRPGRT 118

RESULT 4
AG1997
ribulose-bisphosphate carboxylase activase (EC 6.3.4.-) [similarity] - Nostoc sp. (strain
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1997
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Katanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1997
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-414 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA877899.1; PID:g17135353; GSPDB:GN02179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: rca

C:Superfamily: ribulose-bisphosphate carboxylase activase
C:Keywords: ATP; ligase; nucleotide binding; P-loop
F:37-44/Region: nucleotide-binding motif A (P-loop)
F:43/Binding site: ATP (Lys) #status predicted

Query Match 44.9%; Score 44.5; DB 2; Length 414;
Best Local Similarity 39.3%; Pred. No. 16;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 NVPG-----HERMGRGRTSSKELA 19
      |||
Db 23 NIFGVRVPLILGHRGKGEKIFQCELA 50

RESULT 5
T34592
DEAD-box RNA helicase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 17-Nov-2000
C:Accession: T34592
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A:Reference number: 221548
A:Accession: T34592
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-498 <SMUR>
A:Cross-references: EMBL:AL021529; PIDN:CAAL6457.1; GSPDB:GN00070; SCOEEDB:SC10A5.25c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEEDB:SC10A5.25c

Query Match 43.9%; Score 43.5; DB 2; Length 498;
Best Local Similarity 55.0%; Pred. No. 29;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 NYPGHERMGRGRT-SSKELA 19
      || : ||||| ||| ||
Db 103 SLAGRDIMGRGRTSGKTLA 122

RESULT 6
F95090
conserved hypothetical protein SP0782 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95090
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.;
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 291, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morr
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95090
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74919.1; PID:g14972257; GSPDB:GN00164; TIGR
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0782

Query Match 43.4%; Score 43; DB 2; Length 79;
Best Local Similarity 53.3%; Pred. No. 5.5;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PCHERMGRGRTSSKE 17
      || : ||||| |||
Db 53 PDHTKMGKGTISNE 67

RESULT 7
H97958
conserved hypothetical protein Spr0630 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: H97958
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.;
O, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S
y, R.; Sun, P.M.; Winkler, M.F.;
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H97958
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99494.1; PID:g15458279; GSPDB:GN00174
C:Genetics:
```





C:Superfamily: lysine decarboxylase cad

Query Match 42.9%; Score 42.5; DB 2; Length 491;  
 Best Local Similarity 55.6%; Pred. No. 41;  
 Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 NYPGHERMGRGRTSSKEL 18  
 :||||:| ||| ||  
 Db 27 DVPGHKR-GRGNPELVEL 43

#### RESULT 13

H97973  
 lysine decarboxylase (EC 4.1.1.18) [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: H97973  
 R:Holkins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; B  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: H97973  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-491 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK99620.1; PID:g15458416; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: cad  
 C:Superfamily: lysine decarboxylase cad  
 C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 42.9%; Score 42.5; DB 2; Length 491;  
 Best Local Similarity 55.6%; Pred. No. 41;  
 Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 NYPGHERMGRGRTSSKEL 18  
 :||||:| ||| ||  
 Db 27 DVPGHKR-GRGNPELVEL 43

#### RESULT 14

I40353  
 probable ribosomal protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T40353  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
 submitted to the EMBL Data Library, March 1997  
 A:Reference number: Z21922  
 A:Accession: T40353  
 A>Status: preliminary; translated from GB/EMBL/DDHU  
 A:Molecule type: DNA  
 A:Residues: 1-326 <WOC>  
 A:Cross-references: EMBL:AL020270; PIDN:CAA17794.1; GSPDB:GN00067; SPTB:SPBC3B9.14C  
 A:Experimental source: strain 972h; cosmid c3B9  
 C:Genetics:  
 A:Gene: SPDB:SPBC3B9.14C  
 A:Map position: 2  
 A:Introns: 188/3; 275/3

Query Match 42.4%; Score 42; DB 2; Length 326;  
 Best Local Similarity 50.0%; Pred. No. 33;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKELA 19  
 ||:|:| ||  
 Db 257 GHLLGGQGASSLNL 272

#### RESULT 15

S49464  
 recombination protein recA - Rhodobacter capsulatus  
 N:Alternate names: recombinase A  
 C:Species: Rhodobacter capsulatus  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 02-Feb-2001  
 C:Accession: S49464  
 R:Fernandez de Henestrosa, A.R.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S49464  
 A:Accession: S49464  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-355 <PER>  
 A:Cross-references: EMBL:X82183; NID:g558630; PIDN:CAA57673.1; PID:g558631  
 C:Genetics:  
 A:Gene: recA  
 C:Superfamily: recombination protein recA  
 C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; nucleotide binding; P-  
 F:78-85/Region: nucleotide-binding motif A (P-loop)  
 F:152-157/Region: nucleotide-binding motif B  
 F:84/Binding site: ATP (Lys) #status predicted  
 Query Match 42.4%; Score 42; DB 2; Length 355;  
 Best Local Similarity 50.0%; Pred. No. 36;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKE 17  
 | ||:| ||:| ||:  
 Db 306 GDERIGQGRENNAKQ 319

Search completed: October 1, 2003, 09:41:26  
 Job time : 42 secs